BRAC UNIVERSITY

Inspiring Excellence

Department of Computer Science and Engineering

Thesis Report

"Pattern Recognition through Brushstroke Analysis: An Inspection to Classify van Gogh from Others"

Souradeep Dasgupta
13101146
Md. Ayaz Hossain
13101126

Supervised by:
Moin Mostakim

Submitted on 18th April, 2017

DECLARATION

We, hereby declare that this is an original report written by us with our findings, and has not been published or presented in parts or as a whole for any other previous degree. Resources and materials by other researchers used as guidelines for our research are carefully mentioned in reference citations.

Signature of the Authors:

…………………………

Souradeep Dasgupta

…………………………

Md. Ayaz Hossain

Signature of the Supervisor:

……………………………

Moin Mostakim

Lecturer

Dept. of Computer Science and Engineering

BRAC University

Art Verification

The images from the data set used to acquire the features were patched and segmented for proper inspection, more accurate data handling, and to minimize the CPU and GPU usage. Those patches and segments have been done under supervision, and have been verified as properly usable data samples by an expert in that field, Mustapha Khalid Palash – who is a painter and a renowned architect of Bangladesh.

The section of art of the whole project has been authenticated by Mustapha Khalid Palash, and has been verified by him.

………………………………..

Mustapha Khalid Palash

Disclaimer

The image set of Vincent van Gogh for our research is taken from various digital archives from the museums including The Van Gogh Museum (Amsterdam, Netherlands), The Kröller-Müller Museum (Otterlo, Netherlands), Museum of Modern Arts (Midtown Manhattan, New York, USA), Yale University Art Gallery (New Haven, Connecticut, USA), Minneapolis Institute of Art (Minneapolis, Minnesota, USA), Harvard Art Museum (Cambridge, Massachusetts, USA), The National Museum (London, UK), and Národní galerie v Praze (Prague, Czech Republic). All those images have their copyrights with the museums they belong to, and none of the image was used for any commercial use by us. The images were used solely for research purposes. Some of the van Gogh paintings and all the non van Gogh paintings used here collected from the data set of the competition "Painter by Numbers" [25] in kaggle.com.

Abstract

What makes a painter unique? Does every painter leave a “fingerprint”? How can we determine whether a painting is an original or a fake? Or how can we determine if two different pictures are painted by the same artist or not?
`

The main objective of this project is to find the unique attributes of a painter, and to use those attributes to find similarities amongst other paintings to determine whether that painting is done by the same painter or not. The process includes to check if a painting is done by that painter or not by looking towards the similarities in the set of individual data collected from paintings of various painters, and the thorough analyzation of those data to find similarities through convenient methods. From the movement of brushstrokes [1] and the thorough mathematical analysis amongst those patterns incorporate many aspects of a painter's unique style. The key idea to attain the goal is by using the concept of how do we see art [2], by using the strokes and artistic patterns used by a painter in his/her works to find matches, and using appropriate algorithms to match the brushstroke patterns in other paintings to determine whether the artwork is done by the artist or not. In this experiment, the main focus was on Vincent van Gogh and his works to analyze his paintings, and to analyze his work patterns with his contemporaries and others.

Table of Contents

Page

List of Tables

Page

List of Figures

Page

"If we knew what it was we were doing, it wouldn't be called 'research,' would it?"

– Albert Einstein

Chapter 1

Introduction

In the era of computation, we can see the improvement in computer vision to detect object and distinguish those through intelligent system. From those pattern recognition method discovered by the experts, the inspection in paintings by famous painters can lead to a conclusion where we could get an idea to detect unique attributes of a painter from his painting techniques. From that early times when Vincent van Gogh became famous, art critiques and analysts observed the highly characteristic brushstroke styles of van Gogh, and have relied on discerning these styles for authenticating and dating his works. Evidence substantiates that the brushstrokes of van Gogh are rhythmic; which is, regularly shaped brushstrokes that are tightly arranged, creating a repetitive and patterned impression [12]. Known for the way he applied paint thickly, van Gogh gives a rich texture to the canvas by leaving each brushstrokes visible as opposed to blending or smoothing them [13]. Although the task of artistic classification is entrusted to human experts, recent advances in machine learning and multimedia feature extraction has made the task easier to automate [14],[35]; notwithstanding these all are on very early experimental phase.

In our work, we analyzed the patterns in van Gogh's by statistically analyzing a massive set of automatically extracted brushstrokes from a list of paintings. Afterwards we compared it to other paintings done by van Gogh to check the accuracy of our analysis, and then we took some sample from his contemporaries and other painters with similar stroke patters visible to naked eye to compare and to distinguish those paintings from van Gogh. There are several paintings included in the test set where other painting that are very similar to van Gogh's works, and we also tried to distinguish those from van Gogh's originals to get a rough idea on how to distinguish them using stylometry in terms of the input data collected from the paintings.

1.1 Problem Definition

The art field is filled with works of different painters that covers a huge market of buyers. Detecting genuine paintings from duplicates or forgeries is currently a high-stakes industry. There have been different methods to detect artists from artworks and to detect real arts from fakes, but the field was mostly based on relying on the discerning eyes and experiences of art experts who dedicated in the work and life of the artist(s) [15]. Currently there are different means to detect an authentic art such as ultraviolet fluorescence [16], infrared reflectography [17], x-ray radiography [18], carbon-14 dating [19], painting sampling [20], canvas weave count [21] etc. The main goal is to create a system by which the unique attribute of the painters can be detected through his/her (in our case, van Gogh's) to identify if there exist anything that can be used to classify his paintings from others; to be precise, a primary method to detect whether a painting contains the attribute to be a van Gogh before being examined through the aforementioned technology for the final verification. Since those technology is not yet the commonest and requires a bulk to examine a single paintings, a method is required to verify if a painting is eligible to be tested through those expensive and time consuming methods.

1.2 Motivation

Machine Learning has created a great impact on how information are retrieved from visual data and how to use that in research purposed with efficiency. The development in the field of machine learning algorithms and data analysis methods have shown accurate and optimized performance and it is improving day by day. The idea of working on paintings to recognize the patterns of an artist and to identify if the artist had provided any “signature” in his/her paintings, such as common brushstrokes, the different ratio of use of different types of strokes in his/her painting etc. We were encouraged to do this because there have not been much work in the field where art and science and technology have merged, and it was an opportunity to do something to the fields we are interested in.

1.3 Objectives

The key objective of this thesis is to create a system by which the unique features of a painter can be found through his/her use of brushstrokes in his/her paintings. A success in those feature extraction will lead to –

- Identify the artist through the features of a painting.
- Classify a painting to and by its school of art.
- Create a system to increase the accuracy in pattern-matching in terms of the attributes in the paintings.
- Determine whether a painting is an original or a forged one by comparing that with the features of the attributes of the original painter.
- Reduce forgery in the art market.

1.4 Report Outline

- Chapter 1 contains the formal instruction of the thesis report.
- Chapter 2 contains the literature review, where the problem and its background are mentioned.
- Chapter 3 contains the steps of works behind this thesis. The results and the analysis are also discussed there.
- Chapter 4 contains the conclusion of the report.

Chapter 2
Literature Review

The continuous development in computer vision algorithms has been enabling increasingly complex tasks of image analysis. However, one of the more challenging tasks for computers is the analysis, evaluation and identification of visual art that include associating a specific painting with a painter, authentication of paintings, classifying a painting by its school of art, and many more.[3] In recent times, several researches have been conducted on brushstroke patterns although it has been limited to some specific artists [8][9], i.e. on Van Gogh in the paper Image Processing for Artists Identification by Richard C. et al.; and it is still on experimental phase. Forged paintings have been spreading all around due to the limitation in digital pattern recognition, thus identifying an authentic work of art from a forgery is a high-stakes industry. Our findings suggest that the application of several edge detection algorithms [4],[5],[6] to detect the edges of the strokes and identify and analyze the brushstrokes with different mathematical models to match the patterns can be useful to ascertain the sui generis attributes.

From the collection of analyzed data through the stroke edges, different attributes of van Gogh were found which helped to detect with a good accuracy level whether a given painting is his or not. Moreover the system managed to detect significant differences in the copied paintings of van Gogh done by other painters. The approach towards the outcome came after following several paths which we got from previous work on the similar field, and the other works have been done using mathematical models and algorithms which we found relevant for getting the trusted results.

2.1 Previous Works

There have been works to determine the artist behind a painting using different methods. One of those have been conducted by Shamir et al. [26], where the researchers worked on the paintings of nine different painter to extract the features of those painter using several different algorithms to distinguish each segments of the paintings, and later compare them with the data set they had. The key advantage they had was to use primary data (images collected directly as RAW) and using different transformation techniques to build a trainer set. However, they used random parts of the image to compare and create a match level which could create less accuracy if the system is applied on full size paintings.

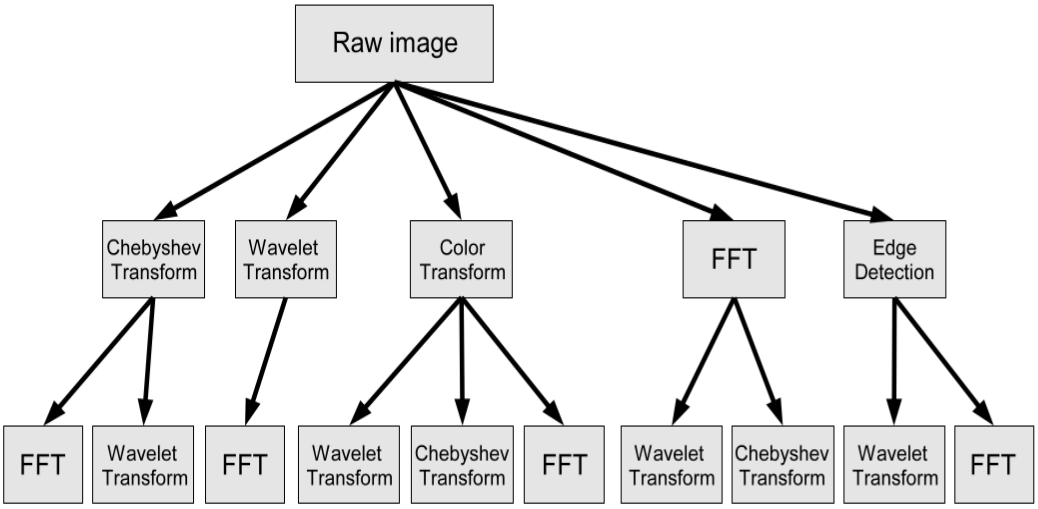


Fig. 1. Image transforms and paths of the compound image transforms.

Another recent work conducted by Shen [27] was an attempt to classify western paintings according to artist. The works are done by using both global color as a feature as well as the texture – mostly local textures. Color features consist of a quantized HSV histogram and a color layout descriptor, while texture features are distinguished by the use of Gabor features. In addition, shape features are incorporated as a histogram by edge detection techniques. The classification is done for 25 classes - artists - by a radial basis function network.

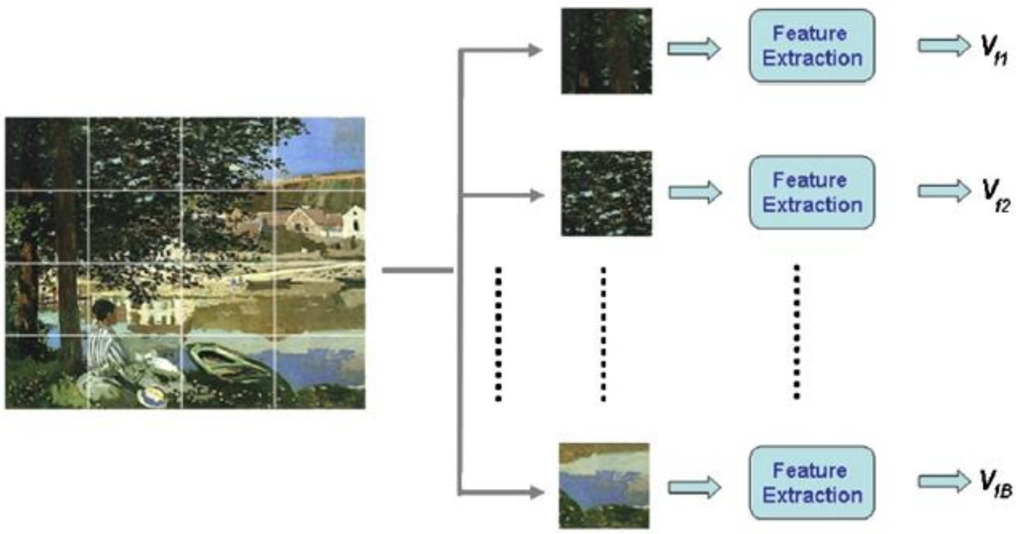


Fig. 2. The procedure of local visual feature extraction.

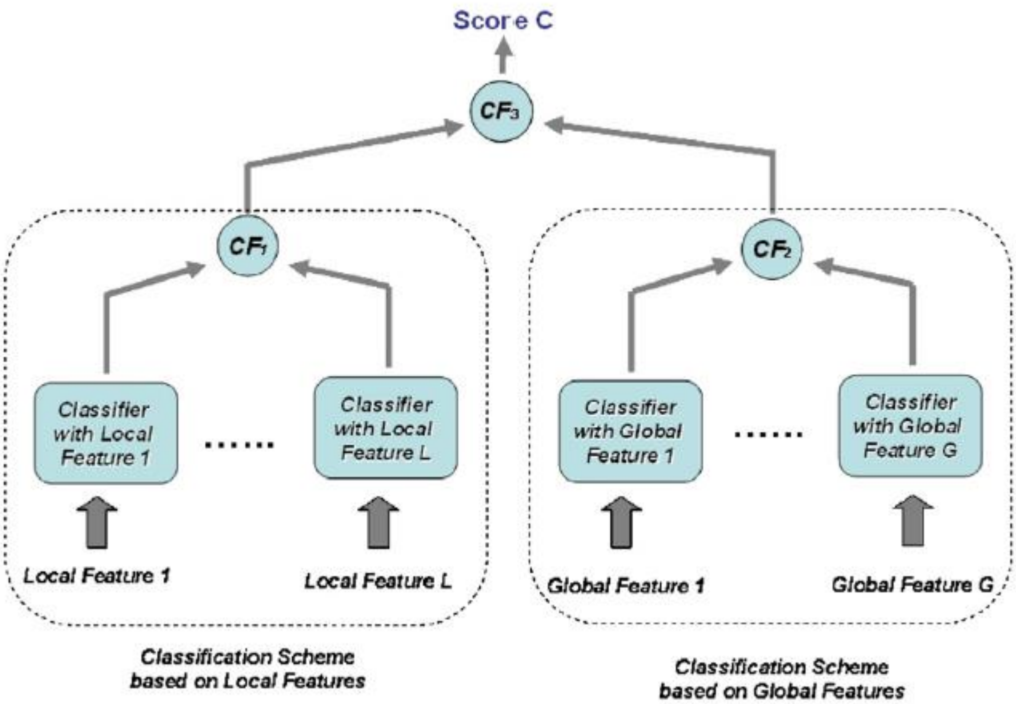


Fig. 3. Classification Scheme on Local and Global Features

We did used a similar pattern to make segmented images and patches like this step, but the image set that we had contains different images of van Gogh that already have lost its original colors due to pigment degradation. The realization that the paintings have been losing their color and detail came very long after van Gogh's death, and during that interval, the painting had not been preserved properly. Thus it did not leave us with the choice of using the color as an attribute, and for the set of van Gogh images, we opted not to go through the aforementioned way of work. So what we believed is that the step could work for the paintings which have been preserved properly, but not suitable for the approach that we have been taking.

An analysis of Vincent van Gogh's painting brushstrokes were done by Johnson, R. et.al., in the paper Image Processing for Artist Identification – Computerized Analysis of Vincent van Gogh's Painting Brushstrokes [1], where the researchers focused on high resolution images of van Gogh's individual brushstrokes to analyze the characteristics of the strokes. The researchers used wavelet transforms to distinguish the pattern is van Gogh’s painting from non van Goghs’.

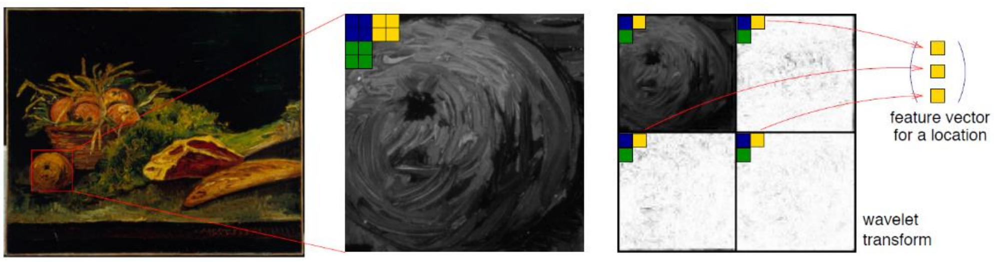


Figure 4: The formation of the texture feature vector based on the wavelet coefficients. The vG painting F219 (Still Life: Basket with Apples, Meat and a Breadroll) is shown.

This approach is one of the most advanced approaches done in the field where the images were properly patched and analyzed. At the same time, the system requires extremely detailed digitized image set, which was not possible for us to get.

The images they have used were provided by the Van Gogh Museum and The Kröller-Müller Museum with files containing high resolution digital photographs of the paintings captured at different spectral wavelengths. Because of the limitation in both the data and the required hardware, it was not feasible for our works.

The project that we followed in the beginning of our project was the works by Li, Gia, et. al., in the paper Rhythmic Brushstrokes Distinguish van Gogh from His Contemporaries: Findings via Automated Brushstroke Extraction [28]. In that paper, the brushstroke extraction was done starting with the edge detection with the help of EDISON system, which we used in our work too.

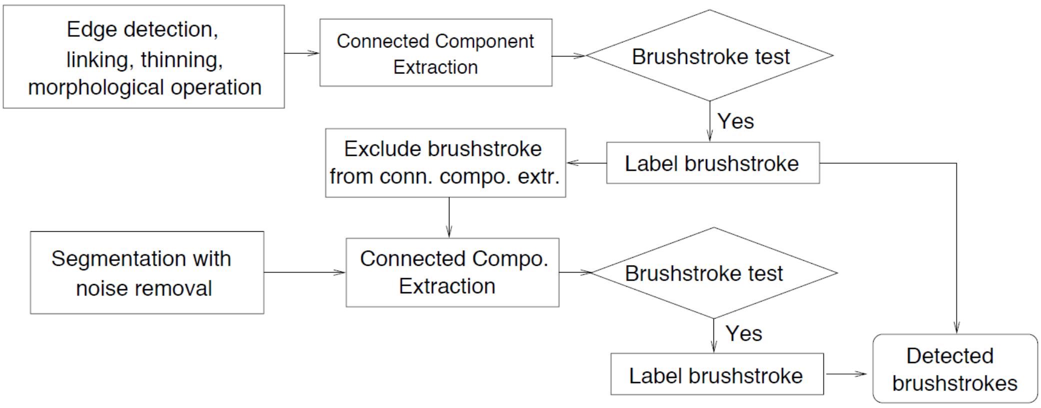
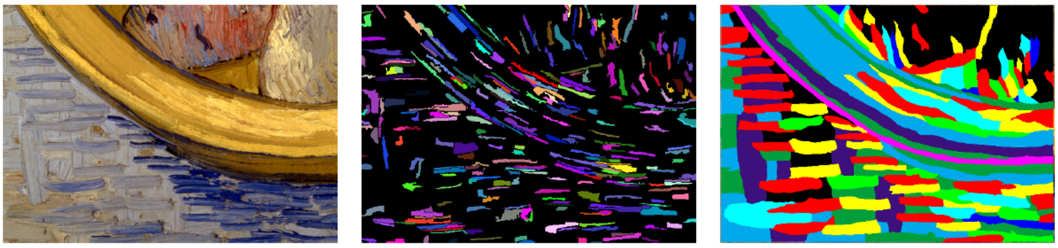


Fig. 5. The flow of the brushstroke extraction algorithm.

They, however used manually marked brushstrokes over those extracted segments, which they did by manually covering those strokes to make a properly visible map – which has accuracy but time consuming, as well as requires a lot of manual efforts.



(a) (b) (c)

Fig. 6. Comparison of automatically extracted and manually marked brushstrokes in example regions. (a) Original image. (b) Automatically extracted brushstrokes. (c) Manually marked brushstrokes.

From all of these paper, we took different ideas to integrate to our works. Since the image processing tools aimed at supplementing the way art historians view on paintings are currently in the earliest stages of development and the necessary data for research has not been made widely available, the researchers are always looking for new methods to imply on their works and to get a proper path to follow to achieve the best result with their limited resources they have. From our observation on the earlier works, we expected that mathematical analysis of a painting's digital representation is possibly the best way to achieve the attributes from a painter's artwork.

2.2 Tools and Approaches

2.2.1 EDISON System

Edge detection is arguably one of the most important operations in low-level computer vision with a plethora of techniques, belonging to several distinct paradigms. For our works, we used EDISON system to detect the brush edges from the canvas. Edge Detection and Image Segmentation (EDISON) System is a low-level feature extraction tool that integrates confidence based edge detection and mean shift based image segmentation. It was developed by the Robust Image Understanding Laboratory at Rutgers University [29].

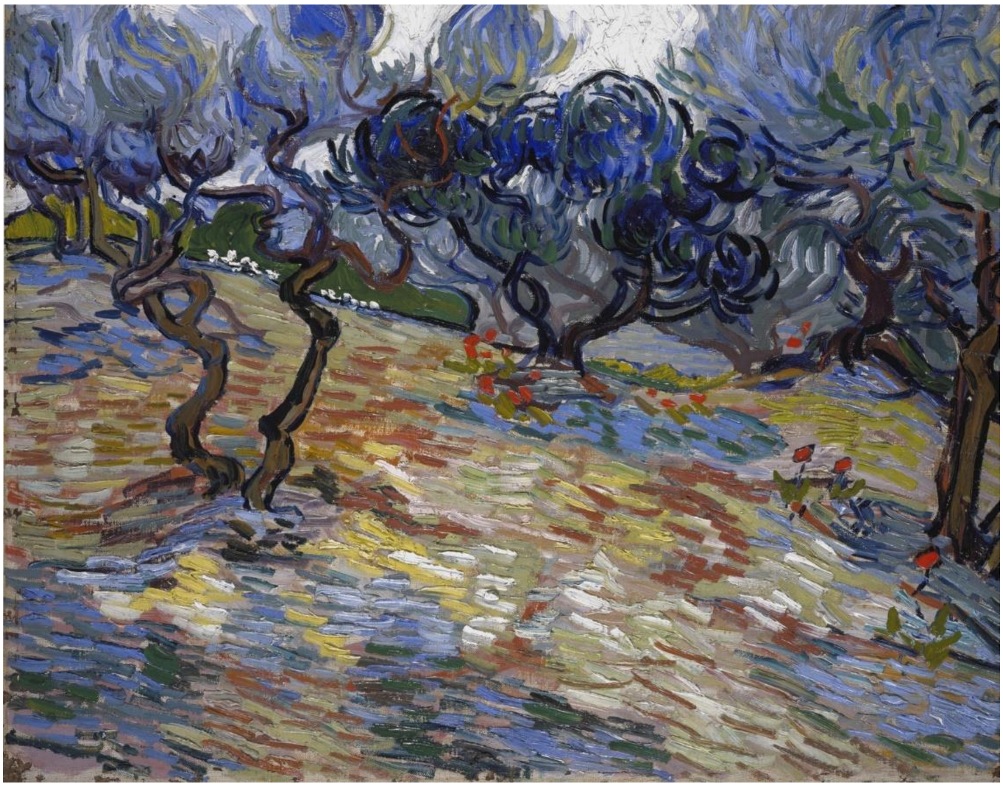


Figure 7: Original Picture Before Edge Detection

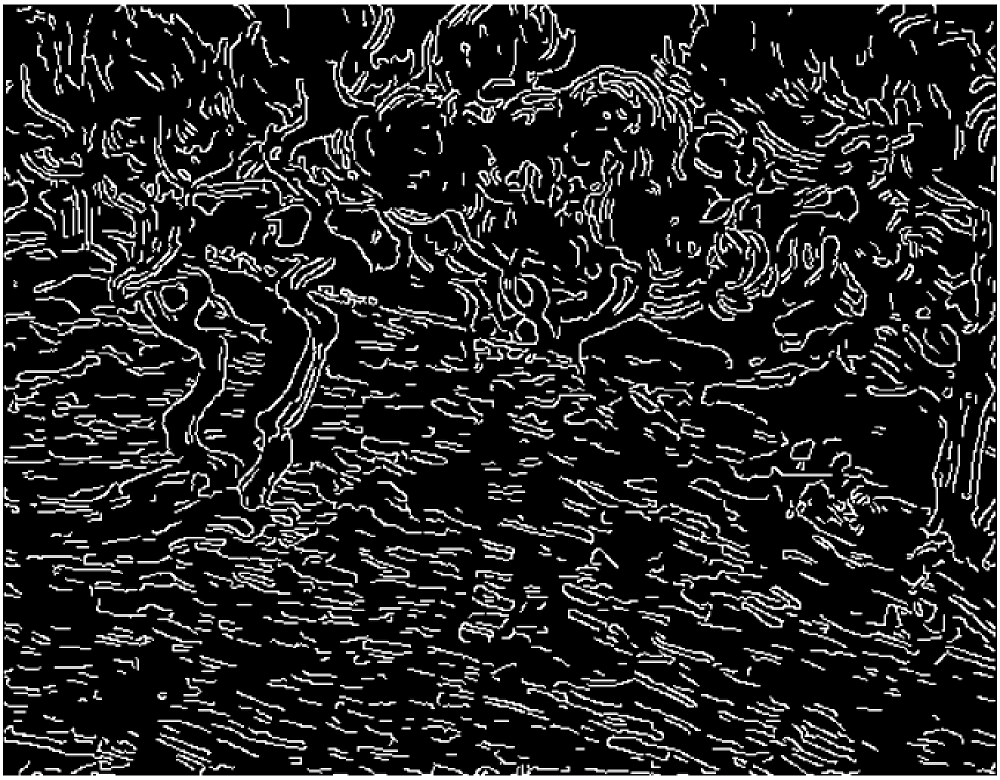


Figure 8: PGM after Edge Detection using EDISON

2.2.2 KNN Algorithm

K-nearest neighbour, or KNN, is a supervised classification algorithm. Because it is supervised, a training set is first used to train the model. The training set consists of vectors in a multidimensional feature space, each with a class label. The training phase of the algorithm consists only of storing the feature vectors and class labels of the training samples.

In the classification phase, *k* is a user-defined constant, and an unlabeled vector (a query or test point) is classified by assigning the label which is most frequent among the *k* training samples nearest to that query point.

To find the nearest points various distance metrics can be used, such as Hamming, Kullback-Leibler, Mahalanobis, Minkowski, or simply Euclidean [31]. Because of the simplicity in our use case, we used Euclidean distance.

The Euclidean distance between the points – p and q is the length of the line segment connecting them. In Cartesian co-ordinates, if p = (p_1, p_2,..., p_n) and q = (q_1, q_2,..., q_n) are two points in Euclidean n-space, then the distance (d) from p to q, or from q to p is given by the Pythagorean formula:

$$\begin{aligned}\mathrm{d}(\mathbf{p},\mathbf{q}) = \mathrm{d}(\mathbf{q},\mathbf{p}) &= \sqrt{(q_1-p_1)^2 + (q_2-p_2)^2 + \cdots + (q_n-p_n)^2} \\ &= \sqrt{\sum_{i=1}^{n}(q_i-p_i)^2}.\end{aligned}$$

To illustrate using KNN, an example is provided below. Here two classes of training set are recorded as blue squares or red triangles. The test sample to be classified is drawn as a green circle.

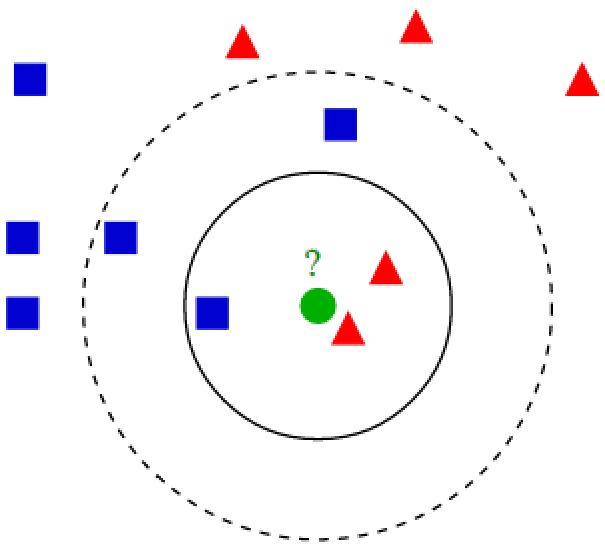


Figure 9: KNN Test Sample

The test sample (green circle) should be classified either to the first class of blue squares or to the second class of red triangles. If k = 3 (solid line circle) it is assigned to the second class because there are 2 triangles and only 1 square inside the inner circle. If k = 5 (dashed line circle) it is assigned to the first class (3 squares vs. 2 triangles inside the outer circle).

2.2.3 Rule of Thirds

The Rule of Thirds is probably one of the most basic rules that has been used in painting for ages. It is a compositional rule of thumb that is commonly used in the visual arts today including painting, photography and design. The Rule of Thirds is actually a guideline more than a rule. It is intended to help the artist with the placement of the elements and focal point within the composition. It was first mentioned by John Thomas Smith in his book *Remarks on Rural Scenary* [30], where he explained –

“Analogous to this "Rule of thirds", (if I may be allowed so to call it) I have presumed to think that, in connecting or in breaking the various lines of a picture, it would likewise be a good rule to do it, in general, by a similar scheme of proportion;

for example, in a design of landscape, to determine the sky at about two-thirds; or else at about one-third, so that the material objects might occupy the other two: Again, two thirds of one element, (as of water) to one third of another element (as of land); and then both together to make but one third of the picture, of which the two other thirds should go for the sky and aerial perspectives. This rule would likewise apply in breaking a length of wall, or any other too great continuation of line that it may be found necessary to break by crossing or hiding it with some other object : In short, in applying this invention, generally speaking, or to any other case, whether of light, shade, form, or color, I have found the ratio of about two thirds to one third, or of one to two, a much better and more harmonizing proportion, than the precise formal half, the too-far-extending four-fifths—and, in short, than any other proportion whatever."

Although artists tend to break this guideline, and van Gogh was no different from them, we can still see the application of rule of thirds in many of his paintings; especially his painting of fields and landscapes. The stroke variation also plays a key role in those paintings which we were focused on to analyze.

Chapter 3

Work and Analysis

We have conducted our analysis in two parts. In each part we extracted and studied different key features of van Gogh's art that distinguish him from other artists. The two parts of our analysis are as follows –

1) Analysis of length and angle deviations of simple strokes
2) Analysis of distribution of discernible strokes.

3.1 Analysis of Length and Angle Deviations of Simple Strokes

Based on the premise that van Gogh's paintings contain tightly packed rhythmic brushstrokes, we analysed the strokes from small randomly cropped patches in his paintings in search of patterns. We studied patches from 12 van Gogh paintings and from 6 non-van Gogh paintings. Each patch was 200x200 pixels.

For each patch, we detected its edges using Edison (see section 2.2.1), which produced the edges as grey scale in a .pgm file. We then identified all strokes from the generated .pgm file. From all the identified strokes the simple strokes were isolated. The length and the angle deviations of the simple strokes were then calculated using the statistical rule for standard deviation. In case of a finite data set x_1, x_2, ..., x_N,, the standard deviation is –

$$\sigma = \sqrt{\frac{1}{N}\left[(x_1-\mu)^2+(x_2-\mu)^2+\cdots+(x_N-\mu)^2\right]}, \text{ where } \mu = \frac{1}{N}(x_1+\cdots+x_N)$$

Finally the calculations of all studied patches were tabulated and was used for the purpose of analysis.

The following flow chart illustrates our workflow for each patch –

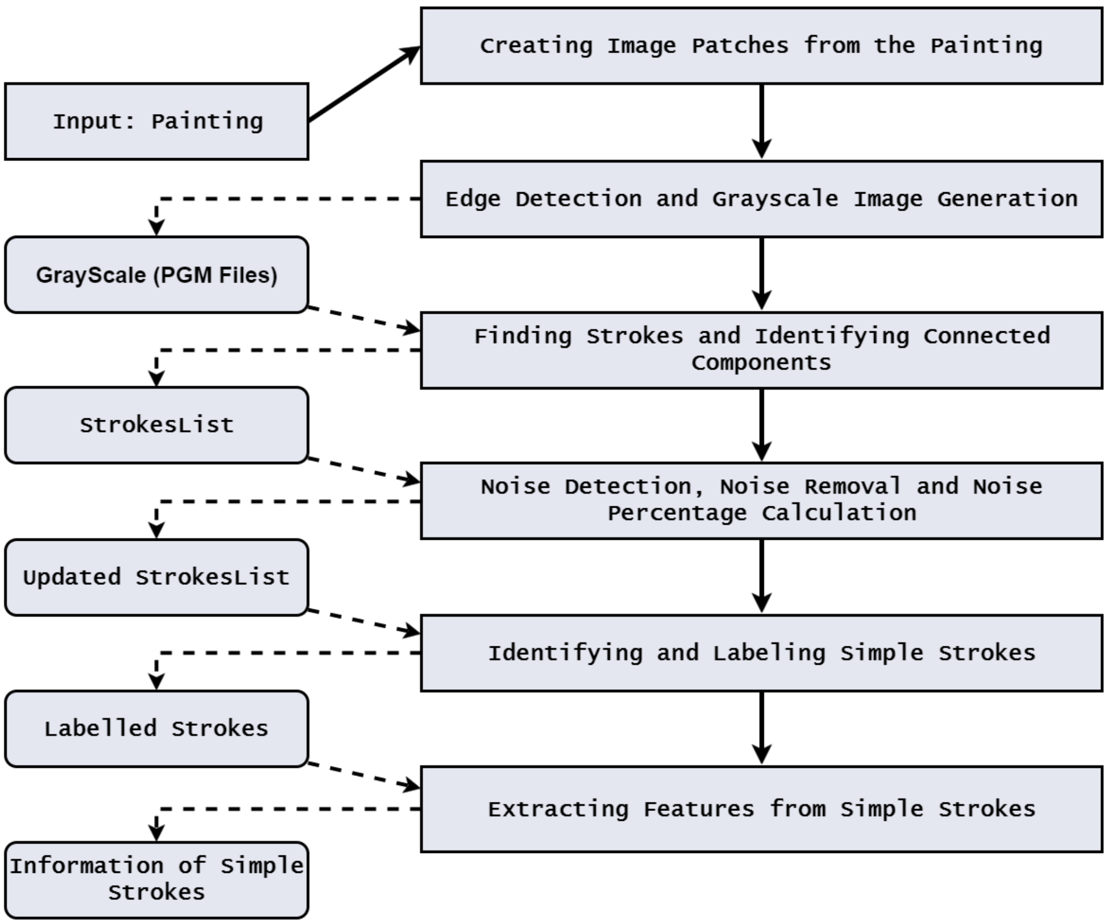


Figure 10: Workflow of Simple Strokes Analysis

3.1.1 Generating Grayscales and Identifying All Strokes

Using Edison, a grey scale representation was created for the patch and stored in a .pgm file. The grey scale image was read using Matlab and stored in 200x200 matrix for further processing.

We used Matlab scripts [Appendix A] to create a graph data structure, represented by a sparse matrix, in which the white pixels were considered nodes. Two nodes were directly connected to each other if they were adjacent. Depth-first search was recursively run on the graph to find the connected components. Each connected component represented a stroke.

3.1.2 Identifying and Isolating Simple Strokes

Any stroke that had less than 10 pixels were discarded as noise, leaving the rest to be checked if they were simple. Each non-discarded stroke was approximated as a straight line starting from its pixel with lowest x and y coordinate values to the pixel with the highest x and y coordinate value. If the difference between the length of the straight line and the actual number of pixels in the stroke was less than 20%, the stroke was classified as simple.

Vincent van Gogh is known for his style of using Impasto – a painting technique where the paint is applied, either by brush or palette knife, thickly. Thus his strokes had identifiable edges, and most of his strokes had linear or uniform motions. Therefore, we are considering those strokes as simple strokes, where the overlapped or zigzag strokes with multiple edges are not being used here. Those strokes considered as simple has been verified by an art expert.

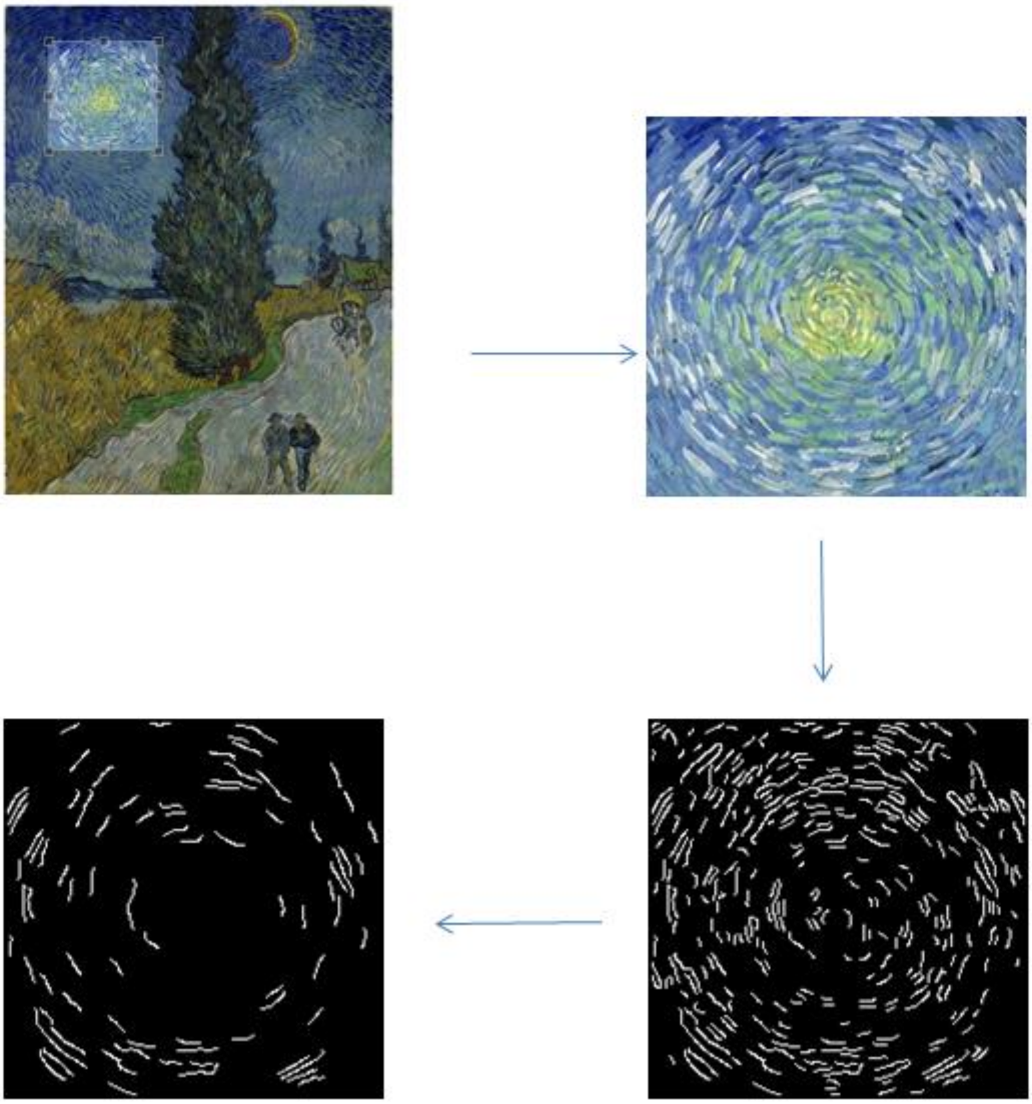


Figure 11: Extracting Features from Patches

3.1.3 Calculating Angle and Length Deviations

We calculated the angles of each stroke from the horizontal using Matlab's built in distance tool. The length of each stroke was taken to be the length of its corresponding approximated straight line. The standard deviations of the angles and lengths were then calculated for the whole set of strokes in a patch.

We were interested in the uniformity of the lengths, but length deviation alone failed to represent this well. To fix the issue we used the ratio of length deviation to mean length in a patch. A low value would represent higher uniformity in lengths and vice versa.

3.1.4 Results

We tabulated average deviations across all patches for each painting. Van Gogh paintings are listed as VG and non-van Gogh ones are list as NVG. Van Gogh and non-Van Gogh paintings are tabulated here separately.

Source Painting	Angle Standard Deviations	Mean Length	Length Standard Deviations	Deviation/Mean
NVG01	69.53004805	23.28491949	14.77517609	0.634538423
NVG02	50.80690986	22.82176172	10.30190939	0.45140728
NVG03	25.62787361	14.83555357	6.876414246	0.463509111
NVG04	69.3675852	21.81749668	12.56045018	0.575705378
NVG05	30.81114616	19.85211551	11.88081923	0.598466155
NVG06	68.56920984	23.10815121	10.1447003	0.439009603

Table 3.1.4.1: Data extracted from Non van Gogh Paintings through the examination on patches

Source Painting	**Angle Standard Deviations**	**Mean Length**	**Length Standard Deviations**	**Deviation/Mean**
VG01	44.30623227	14.18445	5.433199	0.383039
VG02	25.93931364	15.04572	5.49428	0.365172
VG03	19.03136742	19.01902	7.158089	0.376365
VG04	20.77510476	15.48852	5.320951	0.343542
VG05	58.62369703	15.76252	5.100981	0.323615
VG06	30.92588382	14.2752	4.089388	0.286468
VG07	40.71490875	14.46092	5.0357	0.348228
VG08	20.67451043	16.18323	5.95941	0.368246
VG09	64.43748186	14.22547	3.851113	0.270719
VG10	32.82002299	13.8988	4.087787	0.294111
VG11	59.54251228	14.16139	4.263246	0.301047
VG12	19.21228959	14.03531	3.707244	0.264137

Table 3.1.4.2: Data extracted from van Gogh Paintings through the examination on patches

No interesting pattern could be observed in terms of angle deviations. However, It can clearly be observed that van Gogh's paintings have a distinct feature of containing simple strokes that are uniform in length when looking at small regions.

3.2 Analysis of Distribution of Discernible Strokes

This analysis was to study how discernible strokes in van Gogh's art are distributed compared to those of other artists. We based our analysis on the fact that artists use a rule of thumb called 'rule of thirds' (see section 2.2.3), when positioning their subjects in their paintings. Vincent van Gogh was unique from other artists because of his different approach on detailing his artworks through his brushstrokes in different regions of his paintings. In the later period of van Gogh (after he was released from the asylum and before his death), he noticablly used more simpler and small distinguishable strokes in his paintings; one of the reasons behind the mass forgery of his works [36]. Howver, in the last period of Vincent van Gogh paintings, the painted subjects and objects seem to be moving; this dynamical style served to transmit his own feelings about a figure or a landscape. The probability distribution function (PDF) of luminance fluctuations in some impassioned van Gogh paintings, painted at times close to periods of prolonged psychotic agitation of this artist, compares notable well with the PDF of the velocity differences in a turbulent flow as predicted by the statistical theory of Kolmogorov [33]. Because van Gogh predominantly used prominent thick strokes to represent features such as the sky, in contrast to smoothing them, we hypothesised that studying how van Gogh's discernible strokes are spread out throughout the rule of thirds regions would reveal a pattern, and thus paintings from the aforemention period were taken here to analyse and to compare to other paintings.

Figure 12: Portion Seperation Using The Rule of Thirds

For this study we analysed 31 van Gogh, and 11 non van Gogh paintings. The other artists used for comparison were Monet, Rembrandt and Gaugin. The images were collected from a kaggle dataset for the competition 'Painters by Numbers' [25].

Sample paintings for this study was first resized to 600x600 pixels, under the supervision of an art expert. The resized paintings were divided into nine equal 200x200 pixel parts [Appendix B], so that each part corresponds to a rule of thirds region. For each part, strokes were detected using Edison. The total amount of strokes as a percentage of the 600x600 frame was calculated. For each of the nine parts, their stroke percentages of the total strokes in the painting was also calculated. Standard deviation values of percentage of strokes across the regions for each painting was finally caculated.

Taking stroke density, and deviations across the nine segments as X and Y axes respectively, we constructed a graph to visualize our findings.

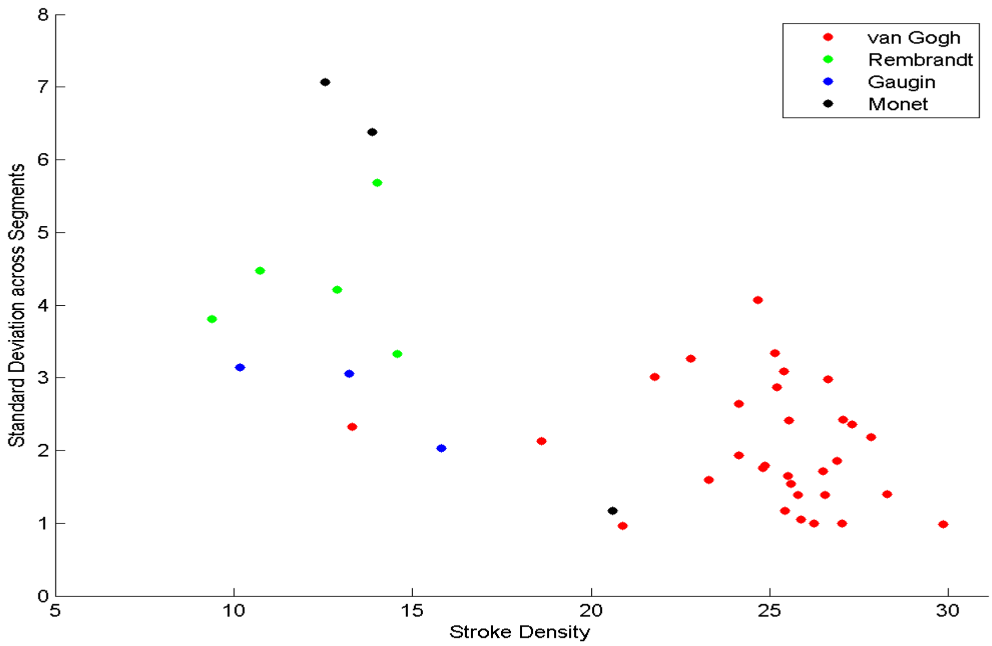


Figure 13: Standard Deviation vs Stroke Density Graph for Trainer

Very interestly, a pattern is noticed. Van Gogh paintings seem to have high stroke densities and low deviations. Similarly different artist's paintings seem to cluster in different areas of the graph, indicating that artist's have signature styles in how they discribute their discernible strokes.

3.2.1 Distinguishing Artists Using Distribution of Strokes with KNN

We used the above data as a model for using K-nearest neighbour (see section 2.2.2) to test if it could distinguish van Gogh from other artists. For our test we chose 3 van Gogh paintings and 3 non-van Gogh paintings. The 'K' parameter for our search was 5. We recorded our results in the tables that follow. We have also plotted points for each test painting graphs to visualize its position with respect to our model. The identifiers of the paintings with NVG represents the non-van Gogh paintings, where VG stands for the paintings of van Gogh.

Painting	Neighbour 1	Neighbour 2	Neighbour 3	Neighbour 4	Neighbour 5
NVG47	Rembrandt	Gaugin	Rembrandt	van Gogh	Gaugin
NVG58	Rembrandt	Monet	Monet	Rembrandt	Rembrandt
NVG61	Rembrandt	Gaugin	Rembrandt	Rembrandt	Gaugin
VG80	van Gogh	van Gogh	van Gogh	van Gogh	van Gogh
VG81	van Gogh	van Gogh	van Gogh	van Gogh	van Gogh
VG93	van Gogh	van Gogh	van Gogh	van Gogh	van Gogh

Table 3.2.1.1: KNN Results for Paintings Compared to Trainer Set

Painting	Neighbour 1	Neighbour 2	Neighbour 3	Neighbour 4	Neighbour 5
NVG47	1.221604994	1.332636448	1.594100472	1.753348547	1.795144983
NVG58	1.145399525	1.221700828	1.439137892	1.457655729	2.445564671
NVG61	2.03507042	2.101206487	3.37536378	4.913406086	4.939305896
VG80	0.245293621	0.348821905	0.522481193	0.564440697	0.666363248
VG81	0.302248027	0.507063256	0.567324752	0.978118395	1.021114089
VG93	0.441557062	0.580763959	0.771509202	1.1337299	1.183601161

Table 3.2.1.2: KNN Results (with values) for Paintings Compared to Trainer Set

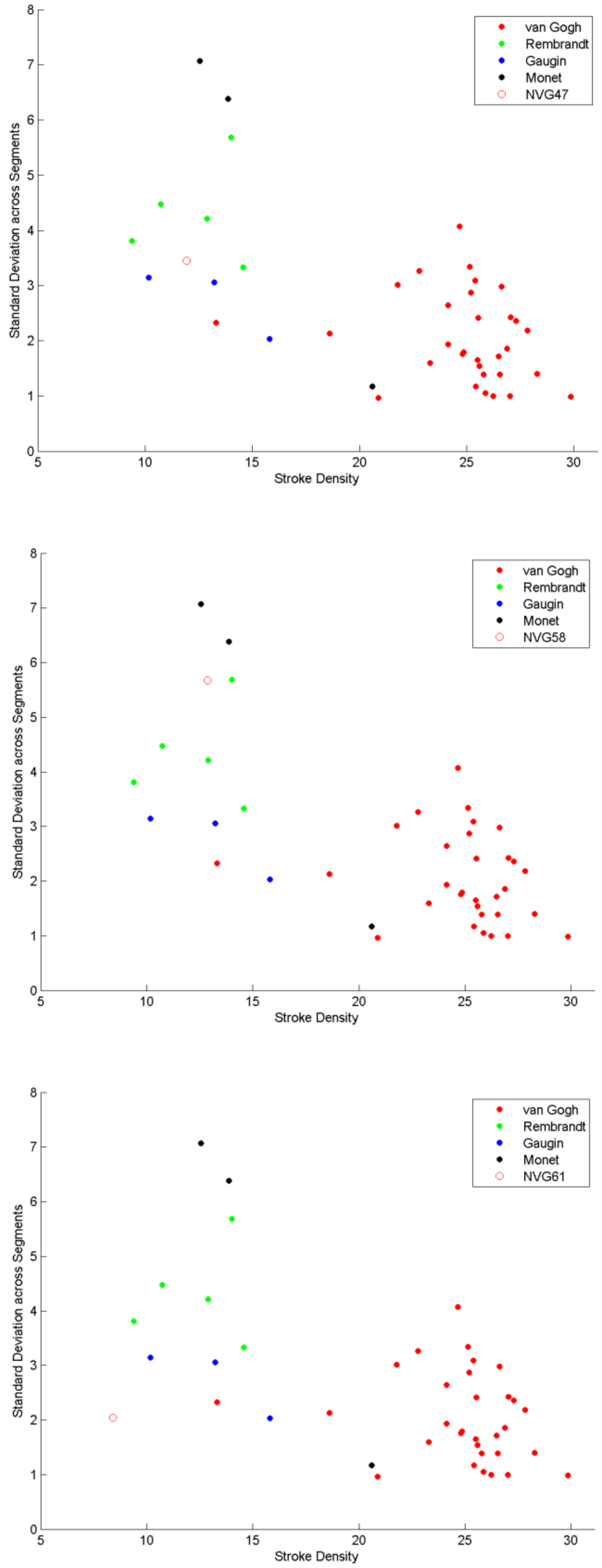
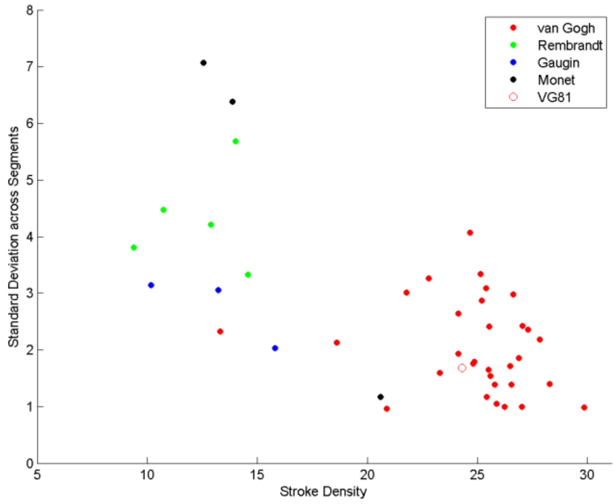
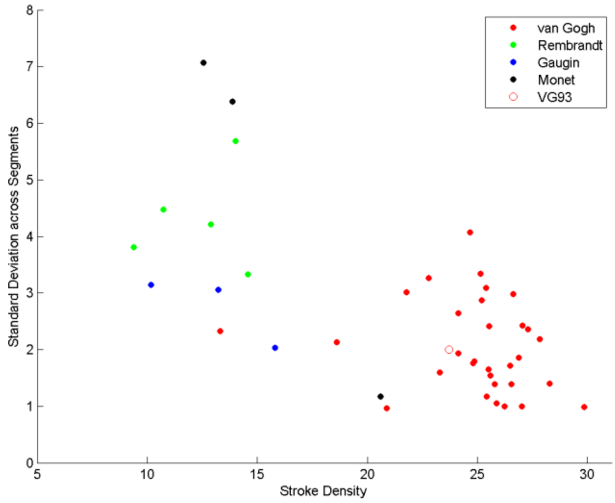


Figure 14, 15, 16: Non van Gogh Painting Comparisons using KNN





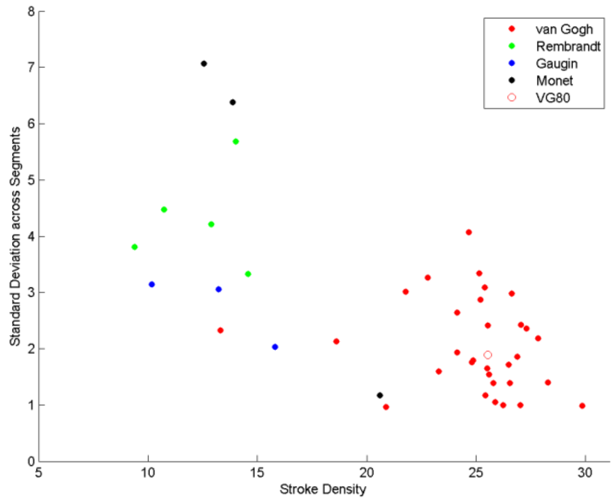


Figure 17, 18, 19: van Gogh Painting Comparisons using KNN

3.3 Limitations

Since the system is still in preliminary development phase, and there have not been much experiments conducted, there were difficulties to follow a guided path; rather we had to focus on retrieving as much attributes as possible. However, the system has limitations such as –

3.3.1 Workable for Basic Analysis

The system is workable for the primary analysis of image forgery detection. As mentioned before in the problem definition section about the different methodologies to detect forgeries, this system works for the primary detection to identify if an image is eligible to go through those steps, as those steps are highly expensive and take a lot of time and resources. This is a step to reduce those expenses and resources, but the system is not a full-functional forgery detector, rather a key step towards the detection process.

3.3.2 Only Applicable for Oil Paintings

The Edge-detecting system mainly works on oil paintings. Due to high levels of polyunsaturated fatty acids, this type of siccative paint always hardens and forms a stable, impermeable film, and creates detectable brush edges when applied. Unless the edges are been flattened intentionally by the painter, the stroke edges can be detected through our system quite easily. Acrylic paints are water-soluble, but become water-resistant when dry, and the dry paints always have the hardness which creates the borders of the strokes visible to the computer vision. On the other hand, water colors are easily dissolvable, thus after the strokes, two or more nearby strokes can dissolve into one another forming a flat overlapping which the system fails to detect.

3.3.3 Mainly Focused on the Works of Post-Impressionist Period

Our work is mainly based on the paintings of Vincent van Gogh, who was a post-impressionist painter. To compare the samples we found, we compared it to classify his works with his contemporaries, as well as the artists who have similar painting techniques. The data set consists of different painters from Impressionist and Post-Impressionist period. Works from popular art movements such as Cubism, Surrealism, and others were not included in our data set.

3.3.4 Failure in Handling Faded Paintings

Numerous paintings of van Gogh are losing their color and details due to aging and chemical effects. Van Gogh was obsessed with the pigments in his paints. He knew the red lake pigments were prone to fading—yet he couldn't resist their vibrancy, and to compensate, van Gogh painted with thick strokes, desperately hoping that the additional paint would keep the colors bright for longer [23]. "Paintings fade like flowers," van Gogh once wrote his brother Theo, "All the more reason to boldly use them too raw, time will only soften them too much." [24] But due to this pigment degradation, those color are fading as well as losing the details in the strokes. To visualize those lost colors and strokes, methods such as macro X-ray fluorescence spectroscopy [22] is needed. Also, paintings with crack were ignored, where there have been works on that [37]. With our limited resources, this was not possible to implement.

3.3.5 Less Data and Computational Resources

The biggest limitation from our side was that we had limited number of high quality images, and we did not have the machines with higher processing power to run the comparison of accuracy. There are different mathematical models to analyse the huge data more efficiently, but that requires better system setup, which we did not have. Application of hyperparameters in an efficient manner could have brought more accuracy in the results, but these parameters would have put huge load on the CPU and the GPU. The workload on the two computers we had took days to be completed which is a lot considering the number of images.

Chapter 4

Conclusion

4.1 Future Works

This thesis is specifically focused on the paintings and the attributes of Vincent van Gogh, so we'd like to work on other painters from different periods and of different art movements too. Our findings suggest that the use of deep neural algorithm [10] can make the system more efficient and more accurate if ample sets of data are provided. Moreover, merging with content-based image retrieval (CBIR) processes such as ArtHistorian [11] can have more benefits in terms of image identification. Moreover, as we did not have a bigger data set, we would be looking for working with more pictures so that we can get better results, and we could look for more attributes of a painter. Since the focus was also on forgery, working with forged paintings could give a better idea on how the mind of a forger works [34], thus we could get more ideas on what to focus on to detect forgeries.

4.2 Conclusion

The system has the ability to detect dissimilarities in the patterns of an artist to a certain level, thus it can work as the primary criteria of a forgery detection tool if it can be trained with more images. Moreover the system can work as a platform for the art enthusiasts to easily learn about the mathematical attributes in a painter's style of painting. This system is an attempt to make a cross-disciplinary interaction of technology and arts to create an intelligent system for both image analysis researchers and art historians.

References

[1] Richard C., et al. Image Processing for Artist Identification. Signal Processing Magazine, IEEE 25.4 (2008): 37-48.

[2] Quiroga R.Q., Pedreira C. How Do We See Art: An Eye-Tracker Study. (2011). 3-6

[3] Shamir, L., Tomasz Macura, Nikita Orlov, D. Mark Eckley, and Ilya G. Goldberg. "Impressionism, Expressionism, Surrealism: Automated Recognition of Painters and Schools of Art " ACM Transactions on Applied Perception 7.2 (2010): 1-17.

[4] "Edge Detection and Image Segmentation (EDISON) System." Edge Detection and Image Segmentation (EDISON) System. Robust Image Understanding Lab, 2002. Web. 07 Apr. 2017. <http://coewww.rutgers.edu/riul/research/code/EDISON/doc/ref.html>.

[5] Kovesi, P. (2001). Edge Detection, Edge Linking, and Line Segment Fitting. N.p., n.d. Web. 7 Apr. 2017. <http://www.peterkovesi.com/matlabfns/LineSegments/example/index.html>.

[6] P. Meer, B. Georgescu: "Edge detection with embedded confidence." IEEE Trans. Pattern Anal. Machine Intell., 23, 1351-1365, December 2001.

[7] Petkov, N., and M.B Wieling. "Gabor Filter for Image Processing and Computer Vision." Web-enabled Image Processing Operators. N.p., n.d. Web. 07 Apr. 2017. <http://matlabserver.cs.rug.nl/cgi-bin/matweb.exe>.

[8] Zujovic J, Gandy L, Friedman S, Pardo B, Pappas TN. "Classifying Paintings by Artistic Genre: an Analysis of Features and Classifiers," Multimedia Signal Processing, 2009. IEEE International Workshop on MMSP '09.

[9] Keren, D., “Painter identification using local features and naive bayes,” in Pattern Recognition, 2002. Proceedings. 16th International Conference on, vol. 2, 2002.

[10] Gatys L.A., Ecker A.S., Bethge M. (2015). A Neural Algorithm of Artistic Style. pp. 9-12

[11] B. Gunsel, S. Sariel, and O. Icoglu, “Content-based access to art paintings,” in Image Processing, 2005. ICIP 2005. IEEE International Conference on, vol. 2, Sept. 2005, pp. 558–561.

[12] Hetebrügge, Jörn. Vincent van Gogh 1853 - 1890. Bath: Parragon, 2008. Print. Art Ser.

[13] (U.S.), National Gallery of Art. “Examining Portraits.” Eye for Art: Focusing on Great Artists and their Works. 2013. pp 61.

[14] Zujovic J, Gandy L, Friedman S, Pardo B, Pappas TN (2009) "Classifying Paintings by Artistic Genre: an Analysis of Features and Classifiers," Multimedia Signal Processing, 2009. IEEE International Workshop on MMSP '09. pp. 1.

[15] Liu, Haixia, et al. (2014). Geometric tight frame based stylometry for art authentication of van Gogh paintings. coRR, abs/1407.0439. pp. 1-2.

[16] Liang, L.H., Apparatus and methods for authentication using partially fluorescent graphic images and OCR characters, US patent 6,373,965, 16 Apr. 2002.

[17] De Boer, J.V.A., A note on the use of an improved infrared Vidicon for reflectography of paintings, Stud. Conserv. 19 (1974) 97–99.

[18] Newman, R., Applications of x rays in art authentication: radiography, x-ray diffraction, and x-ray fluorescence, in: Pho-tonics West’98 Electronic Imaging, International Society for Optics and Photonics, 1998, pp.31–41.

[19] “Carbon-14 Dating.” Dating of Painting. N.p., n.d. Web. 10 Apr. 2017. <https://www.artexpertswebsite.com/authentication/materials-dating.php>.

[20] Caddy, B., Forensic Examination of Glass and Paint: Analysis and Interpretation, CRC Press, 2002.

[21] Berger, G.A., Weave interference in vacuum lining of pictures, Stud. Conserv. 11 (1966) 170–180.

[22] Dik, Joris, et al. "Visualization of a Lost Painting by Vincent van Gogh Using Synchrotron Radiation Based X-ray Fluorescence Elemental Mapping." Analytical Chemistry 80.16 (2008): 6437-6440.

[23] Everts, Sarah. "Van Gogh's Fading Colors Inspire Scientific Inquiry." Chemical & Engineering News (C&EN) 1 Feb. 2016: 32-33. Print.

[24] Vincent van Gogh – The Letters. N.p., n.d. Web. 10 Apr. 2017. <http://vangoghletters.org/vg/letters/let595/translation.html>.

[25] "Painter by Numbers | Kaggle." Painter by Numbers | Kaggle. N.p., n.d. Web 11 Apr. 2017. <https://www.kaggle.com/c/painter-by-numbers/data>.

[26] Shamir, Lior, Tomasz Macura, Nikita Orlov, D. Mark Eckley, and Ilya G. Goldberg. "Impressionism, expressionism, surrealism." ACM Transactions on Applied Perception 7.2 (2010): 1-17.

[27] Shen, J., "Stochastic modeling western paintings for effective classification," Pattern Recognition, vol. 42, no. 2, pp. 293 – 301, 2009, learning Semantics from Multimedia Content.

[28] Li, Jia, Lei Yao, E. Hendriks, and J. Z. Wang. "Rhythmic Brushstrokes Distinguish van Gogh from His Contemporaries: Findings via Automated Brushstroke Extraction." IEEE Transactions on Pattern Analysis and Machine Intelligence 34.6 (2012): 1159-176.

[29] "Robust Image Understanding Laboratory." *Robust Image Understanding Laboratory*. N.p., n.d. Web. 16 Apr. 2017. <http://coewww.rutgers.edu/riul/>.

[30] Smith, John Thomas. Remarks on rural scenery: with twenty etchings of cottages, from nature; and some observations and precepts relative to the pictoresque. USA: ECCO Print Editions, 2012. pp. 17.

[31] Walters-Williams J., Li Y. Comparative Study of Distance Functions for Nearest Neighbors. In: Elleithy K. (eds) Advanced Techniques in Computing Sciences and Software Engineering. (2010). Springer, Dordrecht.

[32] Gurney, James. "Texture and Impasto." Imaginative Realism: How to Paint What Doesn't Exist. N.p.: Andrews McMeel Publishing, 2009. 194-95.

[33] Aragón, J. L., Gerardo G. Naumis, M. Bai, M. Torres, and P. K. Maini. "Turbulent Luminance in Impassioned van Gogh Paintings." Journal of Mathematical Imaging and Vision 30.3 (2008): 275-83

[34] Charney, Noah. The Art of Forgery: the minds, motives and methods of master forgers. London: Phaidon, 2015.

[35] Hertzmann, Aaron. Algorithms for Rendering in Artistic Styles. Ph.D. thesis. New York University. (2001): 19-32.

[36] Feilchenfeldt, Walter. "Van Gogh Fakes: The Wacker Affair, with an Illustrated Catalogue of the Forgeries." Simiolus: Netherlands Quarterly for the History of Art 19.4 (1989): 289-300.

[37] Abas, Fazly S., and Kirk Martinez. "Classification of painting cracks for content-based analysis." Machine Vision Applications in Industrial Inspection XI (2003): 1-12.

Appendix

Appendix A – Stroke Identification using MATLAB

```
%for each image read into matrix 'A' do the following -
[ row, column ] = find(A);
adjacency = createAdjacency( row, column );

function adjacency = createAdjacency( row, column )
%create an sparse adjacency matrix representing a graph of where each
%node lies in specified rows and columns in row and column arrays
%respectively

noOfElements = length(row); %find number of nodes
adjacency = sparse(noOfElements, noOfElements);

%for each adjacent 8 areas, check if there is a node. If there is add
%the connection to the sparse matrix by assigning value of 1
%appropriately.
for a = 1 : noOfElements
    %up-left
    foundElement = findElement(row, column, row(a) - 1, column(a) -
1);
    if foundElement
        adjacency(a, foundElement) = 1;
    end

    %up
    foundElement = findElement(row, column, row(a) - 1, column(a));
    if foundElement
        adjacency(a, foundElement) = 1;
    end

    %up-right
    foundElement = findElement(row, column, row(a) - 1, column(a) +
1);
    if foundElement
        adjacency(a, foundElement) = 1;
    end

    %right
    foundElement = findElement(row, column, row(a), column(a) + 1);
    if foundElement
        adjacency(a, foundElement) = 1;
    end

    %down-right
    foundElement = findElement(row, column, row(a) + 1, column(a) +
1);
    if foundElement
        adjacency(a, foundElement) = 1;
    end
```

```
    %down
    foundElement = findElement(row, column, row(a) + 1, column(a));
    if foundElement
        adjacency(a, foundElement) = 1;
    end

    %down-left
    foundElement = findElement(row, column, row(a) + 1, column(a) -
1);
    if foundElement
        adjacency(a, foundElement) = 1;
    end

    %left
    foundElement = findElement(row, column, row(a), column(a) - 1);
    if foundElement
        adjacency(a, foundElement) = 1;
    end
end

end


function x = findElement( row, column, atRow, atColumn )
% check if a node exists at the specified atRow and atColumn indexes,
%and return it in x. Node positions are listed in the 'row' and
%'column' arrays.

x = 0;
found = 0;
if (atRow > 0 && atRow <= 200 && atColumn > 0 && atColumn <= 200)
%out %of range
    elemsAtRow = find(row == atRow); % get list of nodes at row
'atRow'
    elemsAtColumn = find(column == atColumn);  % get list of nodes at
%column 'atColumn'

% check if any node in elemsAtRow is the same as in elemsAtColumn, if
%it exists, assign it to x and break.
    for a = 1:length(elemsAtRow)
        for b = 1:length(elemsAtColumn)
            found = (elemsAtRow(a) == elemsAtColumn(b));
            if found
                x = elemsAtRow(a);
                break
            end
        end
        if found
            break
        end
    end
end

end
```

Appendix B – Java code snippet to create nine equal segments following rule of thirds, using Java ImageIO

```java
import javax.imageio.ImageIO;

import java.awt.image.BufferedImage;

import java.io.*;

import java.awt.*;


public class ImageSplitTest {

    public static void main(String[] args) throws IOException {


        File file = new File("vg001.jpg"); //van Gogh sample #1 taken

        FileInputStream fis = new FileInputStream(file);

        BufferedImage image = ImageIO.read(fis);


//divide into3 equal rows and columns

        int rows = 3;

        int cols = 3;

        int chunks = rows * cols;


        int chunkWidth = image.getWidth() / cols; // determines the chunk
                                                 // width and height

        int chunkHeight = image.getHeight() / rows;

        int count = 0;

        BufferedImage imgs[] = new BufferedImage[chunks];//Image array to
                                                        //hold image chunks

        for (int x = 0; x < rows; x++) {

            for (int y = 0; y < cols; y++) {

                //Initialize the image array with image chunks

                imgs[count] = new BufferedImage(chunkWidth,
                                  chunkHeight, image.getType());


                // draws the image chunk

                Graphics2D gr = imgs[count++].createGraphics();
```

```
                gr.drawImage(image, 0, 0, chunkWidth, chunkHeight,
                              chunkWidth * y, chunkHeight * x,
                              chunkWidth * y + chunkWidth,
                              chunkHeight * x + chunkHeight, null);

                gr.dispose();

            }

        }

        System.out.println("Splitting done");


        //writing mini images into image files

        for (int i = 0; i < imgs.length; i++) {

            ImageIO.write(imgs[i], "jpg", new File("img" + i +
                         ".jpg"));

        }

        System.out.println("Mini images created");

    }

}
```